(Effect) Size Matters:

And So Does the Calculation

Melissa L. Olive & Jessica H. Franco

The University of Texas at Austin

Author Contact:
Melissa L. Olive, Ph.D., BCBA
The University of Texas at Austin
Department of Special Education
1 University Station D5300
Austin, TX 78712
(512) 475-6585
molive@mail.utexas.edu

Abstract

The purpose of this paper is to present the types of measures that may be used to describe intervention effects from single subject designs. A regression approach and several non-regression approaches are described. Non-regression approaches include Standard Mean Difference, Percentage of Non-Overlapping Data, Percent Reduction, and Percentage Exceeding the Median. Researchers are encouraged to combine a non-regression measure along with considerations of methodological rigor and visual analysis to fully appreciate the contributions of single subject intervention data.

(Effect) Size Matters:

And So Does the Calculation

In 2001, the American Psychological Association (APA) noted in its publication manual that effect size calculations should be included in manuscripts submitted for publication. However, researchers utilizing single subject designs have not typically embraced the approach of any analyses beyond that of the traditional visual analysis (Marascuilo & Busk, 1988; Parsonson & Baer, 1977).

In visual analysis of single subject data, researchers have examined data for three changes in the data: trend, variability, and level. Using trend analysis, researchers have examined the direction of the data for an increasing (i.e., upward) or decreasing (i.e., downward) trend. Researchers have also inspected for change in data variability or bounce. Finally, researchers have noted changes in level or mean performance.

Recent trends in the field of education have resulted in an increased need to synthesize data sets from single subject studies. For example, the No Child Left Behind Act (NCLBA; 2001) brought considerable attention to the term evidence-based practice. As Odom and colleagues described (2005), some have claimed that only randomized experimental group designs are appropriate for demonstrating scientific evidence. This precluded single subject studies from being included in contributions of scientific evidence on effective intervention methods. However, others have noted that rigorous single subject research has much to contribute when determining scientific knowledge within the field (Horner, et al., 2005). In order to support the use of single subject research as evidence-based, a process of synthesizing single subject data is needed. Additionally, the Individuals with Disabilities Education Act (2004) mandated that teachers use strategies based on evidence based research. It would be tragic for teachers to utilize only teaching strategies proven with group design research; hence a second need to summarize data from single subject studies. Finally, researchers conducting meta-analyses or research syntheses have needed a method for interpreting and comparing intervention effectiveness of single subject studies. Researchers

and practitioners in the field have tried to synthesize intervention research and effect sizes have been calculated on single subject data (e.g., Ma, 2006; Parker, Hagan-Burke, & Vannest, 2007; Wanzek, et al., 2006).Therefore, the purpose of this paper is to present the types of measures that may be used to describe intervention effects of single subject research designs. Strengths and limitations of each method will be described. Finally, a recommendation will be made to assist in determining which method should be used with which types of single subject data.

Regression Approaches

Allison and Gorman described the use of regression models to calculate effect sizes with single subject data (Allison & Gorman, 1993; Faith, Allison, & Gorman, 1996). In doing so, the dependent measure in the study (e.g., reading fluency or out of seat behavior) served as the dependent measure in the analysis while the intervention sessions serve as the independent variable. A separate regression equation was then obtained for the baseline and intervention data resulting in two regression equations. Finally, the intervention was subtracted from the baseline and divided by the standard deviation of baseline (Hershberger, Wallace, Green, & Marquis, 1999).

It should be noted that data portrayed in single subject graphs are not independent of one another. Often in single subject research, experimenters visually analyze intervention data following each intervention session. This visual analysis might result in modifications to intervention procedures during the subsequent session resulting in data that are dependent on preceding data. For example, if a child was being taught to exchange a graphic symbol for a preferred item, the dependent variable might be rate of independent exchanges. If, on intervention session four, the child exchanged the symbol at a rate of .2 the experimenter might modify the following session by using a different reinforcer with hopes of increasing the exchange rate. This practice would result in data that are serially dependent. Therefore, regression analyses should be avoided with single subject data.

Non-Regression Approaches

Non-regression analyses, however, may be more appropriate for use with single subject data. A variety of non-regression approaches have been described in the literature. These approaches have manipulated the single subject data resulting in values that quantify the degree of intervention effectiveness above and beyond the traditional approach of visual analysis. Each of these approaches has produced a quantifiable value that must be interpreted.

Percentage of Non-Overlapping Data

A widely used non-regression approach has been Percentage of Non-Overlapping Data (PND; Scruggs & Mastropieri, 2001). This calculation has been described as a “meaningful index of treatment effectives” (p 241). To calculate PND, the percentage of data points during intervention that surpassed the extreme values in pretreatment or baseline was calculated. Specifically, in an intervention to increase the dependent variable, the proportion of treatment data points that exceeded the highest baseline value was calculated. During behavior reduction interventions, the proportion of intervention data points that fell below the lowest baseline was calculated. In either case, the number of non-overlapping intervention points was divided by the total number of intervention data points to determine the PND.

Scruggs and Mastropieri (1998) made special recommendations for using PND with specific types of single subject studies. For example, if a return to baseline design was utilized, the first baseline data set should be used. If multiple treatments were tested, the final phase of intervention data should be used.

Scruggs and Mastropieri also provided suggestions for interpreting PND results (1998). They suggested that PND scores above 90 represented very effective treatments, scores from 70 to 90 represented effective treatments, scores from 50 to 70 were questionable, and scores below 50 were ineffective.

One advantage of the PND score has been that behavioral researchers may be able to readily interpret the data. With extensive practice using visual analysis, behavioral

researchers have understood the meaning of 90% of intervention data not overlapping with baseline data. However, this advantage to behavioral researchers might also serve as a disadvantage to non-behavioral researchers who do not understand single subject research designs. Specifically, a reader without extensive experience with visual analysis would most likely lack an understanding of what 90% of non-overlapping data means.

A second disadvantage of PND is that some studies were not appropriate for the calculation. Specifically, Scruggs and Mastropieri (1998) advised that a PND should not be calculated when a data point in baseline is at the ceiling or floor. Specifically, in a behavior reduction study, if one data point in baseline was zero, then PND would automatically be 0% regardless of the number of data points at zero during intervention.

Percent Reduction

Campbell (2000) termed mean baseline reduction (MBR) using procedures originally described by O'Brien and Repp (1990). In this calculation, the mean baseline and mean intervention measurements were determined for the last three sessions of each. The mean of intervention was subtracted from the mean of baseline and divided by the mean of baseline and multiplying by 100. This produced a mean percent reduction from baseline.

This approach has been helpful in determining how much a behavior has decreased during intervention; however, it has lacked usefulness for determining an effect for interventions that increase behavior, particularly when baseline rates of the behavior are zero.

Percentage Exceeding the Median

A relatively new approach has been the percentage of data points exceeding the median of the baseline phase (PEM; Ma, 2006). For intervention studies focusing on increasing behaviors, Ma suggested that reviewers draw a median line for the baseline data and calculate the percentage of data points in intervention that fall above the median line For behavior reduction studies, the percentage of data points below the median line should be calculated.

Several strengths could be found in the PEM approach. First, there have been no reports of situations where PEM could not be used. Second, PEM has been shown to be correlated with author judgments of intervention effectiveness (Ma, 2006). However, as with PND, the meaning of the percentage calculated may be misconstrued by researchers unfamiliar with single subject design. Finally, as Ma reported, this measure failed to show sensitivity to the magnitude of intervention data points above the median line.

Standard Mean Difference

The standard mean difference is one gauge of intervention effectiveness. Busk and Serlin (1992) presented the standard mean difference (SMD) equation. First, the mean difference from baseline to intervention is calculated. Next a standard is calculated. Many times, the standard deviation of baseline serves as that standard. Finally, the difference is divided by the standard. What results is in an actual effect size value (d) that may be more easily understood by readers. Effect sizes should be interpreted as follows: $d = 0.2$ small, $d = 0.5$ medium, and $d = 0.8$ large (Cohen, 1988).

SMD may be calculated in two ways; SMD_{all} and SMD_3. In SMD_{all}, all baseline and all intervention data points are utilized whereas in SMD_3, only the last three data points of baseline and intervention are used. Using only the last 3 data points of baseline and intervention may increase the effect size because, in single subject studies, the last few sessions are usually the best. However, if all the data points are utilized in the calculation, the variability of the data would be captured in the analysis (although not reflected in the actual results). Therefore readers should recognize that SMD_3 results may be inflated and that SMD_{all} results are most likely more accurate.

Olive and Smith (2005) noted that some rules should be established to create standards for calculating SMD. For example, with a reversal design, the original baseline data and the last intervention data should be used. With an alternating treatments design, the superior treatment data should be used. If a multiple baseline design was employed, an effect should be calculated for each person, setting, or behavior in the study. Finally, in a

changing criterion design, the original baseline data and the last intervention data should be used.

The SMD approach offers several strengths. First, average data are used resulting in a formula that may be used in all studies whether the data are increasing in nature (e.g, skill acquisition) or decreasing (e.g., challenging behavior). In this approach, no data need to be discarded due to factors such as overlapping data. The SMD calculation results in an actual *d* score making it more interpretable by readers. Results from other approaches must be interpreted (e.g., is 80% a good effect or an acceptable effect?). Finally, the SMD calculation is simple. Data stored in any spread sheet typically used for graphing may be used without the need for recalculations or re-entry.

Recommendations and Conclusions

Of the non-regression measures, it appears that SMD_{all} may be the most appropriate to use to compare intervention effects during literature reviews or syntheses. Additionally, all data points should be used in the calculation to more accurately describe the true intervention effect and to reduce the likelihood of an inflated effect size. Moreover, the SMD method results in an actual effect size value (*d*) that may be more easily understood than the numbers obtained from calculations of PND or MBR.

It should be noted that all of the non-regressive approaches merely describe changes in the levels of the single subject data. None of the approaches capture the trend or the variability of the data. On the other hand, visual analyses capture all three of these effects. Therefore, non-regression approaches should never be used in lieu of visual analysis, but rather they should be paired with a visual analysis to ensure a comprehensive understanding of the intervention effect.

Moreover, these approaches do not consider the methodological rigor of the study. Horner and colleagues described the quality indicators of methodological rigorous single subject studies (2004). First, they noted that participants and settings should be clearly and operationally described. They noted that it was insufficient to generally describe participant

characteristics. Horner and colleagues stressed the importance of operationally defining the dependent variable. The dependent variable should be measured repeatedly and frequently and authors should report a measure of inter-observer agreement on the dependent variable. Horner and colleagues also described the importance of carefully describing the independent variable and presenting data on procedural fidelity. Finally, Horner and colleagues stressed the importance of demonstrating a functional relationship between the independent variable and change to the dependent variable. They noted that a baseline condition was required and that a minimum of three demonstrations of experimental control were necessary.

In summary, researchers are encouraged to combine considerations of methodological rigor and visual analysis with a non-regression measure in order to fully appreciate the contributions of the single subject intervention data. The most appropriate measure depends on the type of research design, the nature of the data collected, and the purpose for the calculation.

References

Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behavior, Research, and Therapy*, *31*, 621–641.

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington,DC: American Psychological Association.

Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research designs and analysis: New directions for psychology and education* (pp. 187–212). Hillsdale,NJ: Lawrence Erlbaum Associates.

Campbell, J. M. (2000). *Efficacy of behavior interventions to reduce problematic behaviors in persons with autism: A quantitative synthesis of single-subject research*. Unpublished doctoral dissertation, University of Memphis, TN.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Faith, M. S., Allison, D. B., & Gorman, B. S. (1996). Meta-analysis of single-case research. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 245–277). Mahwah,NJ: Lawrence Erlbaum Associates.

Hershberger, S. L., Wallace, D. D., Green, S. B., & Marquis, J. G. (1999). Meta-analysis of single-case designs. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 109–132). Newbury Park,CA: Sage.

Horner, R. H., Carr, E. G., Halle, J., McGee, G, Odom, S. L., & Wolery, M. (2005). The use of single subject research to identify evidenced-based practice in special education. *Exceptional Children, 37*,165-179.

Individuals with Disabilities Education Improvement Act of 2004, 20 U.S.C. § 1400 *et seq.* (2004) (reauthorization of the Individuals with Disabilities Education Act of 1990).

Ma, H. H. (2006). An alternative method for quantitative synthesis of single-subject researches. *Behavior Modification, 30*, 598-617.

Marascuilo, L. A., & Busk, P. L. (1988). Combining statistics for multiple-baseline AB and replicated ABAB designs across subjects. *Behavioral Assessment*, *10*, 1–28.

No Child Left Behind Act of 2001, 20 U.S.C. 70 § 6301 *et seq.* (2002).

O'Brien, S., & Repp, A. C. (1990). Reinforcement-based reductive procedures: A review of 20 years of their use with persons with severe or profound retardation. *Journal of the Association for Persons with Severe Handicaps*, *15*, 148–159.

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children, 37*, 137-148.

Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of All Non-Overlapping Data (PAND): An Alternative to PND. *Journal of Special Education, 40*, 194-204.

Parsonson, B. S., & Baer, D. M. (1978). The analysis and presentation of graphic data. In T. R. Kratochwill (Ed.), *Single subject research: Strategies for evaluating change* (pp. 101–166). New York: Academic.

Scruggs, T. E., & Mastropieri, M. A. (1998). Synthesizing single subject studies: Issues and applications. *Behavior Modification*, *22*, 221–242.

Scruggs, T. E., & Mastropieri, M. A. (2001). How to summarize single participant research: Ideas and applications. *Exceptionality*, *9*, 227–244.

Wanzek, J., Vaughn, S., Wexler, J., Swanson, E. A., Edmonds, M., & Kim, A. H. (2006). A synthesis of spelling and reading interventions and their effects on the spelling outcomes of students with LD. *Journal of Learning Disabilities, 6*, 528-543.